

融合自注意力与渐进抽取机制的电商多场景多任务预测方法研究

王英万¹, 于丽娅^{1*}, 李少波², 李传江³, 徐兆¹

¹贵州大学机械工程学院, 贵州 贵阳

²贵州理工学院机械工程学院, 贵州 贵阳

³贵州大学省部共建公共大数据国家重点实验室, 贵州 贵阳

收稿日期: 2026年4月9日; 录用日期: 2026年4月23日; 发布日期: 2026年6月30日

摘要

随着电商平台由单场景、单目标逐步迈向多场景、多任务的统一建模范式, 如何在充分共享跨场景共性知识的同时有效规避任务间冲突与负迁移, 已成为制约模型可用性与训练稳定性的核心挑战。针对该问题, 提出一种面向电商多场景多任务预测的分层专家增强框架: 以Star场景融合底座为基础, 在嵌入层之后引入自注意力模块, 以显式刻画高阶特征依赖; 同时采用渐进分层抽取替代传统单层MoE的共享机制, 使共享-特化在不同语义层级实现自适应迁移与有效解耦, 从结构上缓解跷跷板现象并抑制负迁移。此外, 模型保留场景/任务专家, 并通过可学习的平衡混合机制增强对场景不均衡与分布漂移的鲁棒性。在两个公开数据集上开展了广泛实验, 结果表明该方法在整体性能与稳定性方面均显著优于对比模型。

关键词

推荐系统, 多场景推荐, 多任务预测, 电子商务, 注意力机制

Research on a Multi-Scenario, Multi-Task Prediction Method for E-Commerce Based on Self-Attention and Progressive Extraction Mechanisms

Yingwan Wang¹, Liya Yu^{1*}, Shaobo Li², Chuanjiang Li³, Zhao Xu¹

¹School of Mechanical Engineering, Guizhou University, Guiyang Guizhou

²School of Mechanical Engineering, Guizhou Institute of Technology, Guiyang Guizhou

³State Key Laboratory of Public Big Data, Guizhou University, Guiyang Guizhou

*通讯作者。

文章引用: 王英万, 于丽娅, 李少波, 李传江, 徐兆. 融合自注意力与渐进抽取机制的电商多场景多任务预测方法研究[J]. 电子商务评论, 2026, 15(6): 1004-1014. DOI: 10.12677/ecl.2026.156720

Abstract

As e-commerce platforms gradually evolve from single-scenario and single-objective frameworks toward a unified multi-scenario and multi-task modeling paradigm, how to fully share cross-scenario common knowledge while effectively avoiding inter-task conflicts and negative transfer has emerged as a core challenge restricting model usability and training stability. To address this issue, we propose a hierarchical expert-enhanced framework for multi-scenario multi-task prediction in e-commerce. Building upon a Star-topology scenario fusion foundation, a self-attention module is introduced after the embedding layer to explicitly capture high-order feature dependencies. Simultaneously, a progressive hierarchical extraction mechanism replaces the traditional single-layer MoE sharing approach. This enables the adaptive transfer and effective decoupling of shared and task-specific representations across different semantic levels, fundamentally mitigating the seesaw phenomenon and suppressing negative transfer from a structural perspective. Furthermore, the model retains scenario and task-specific experts, utilizing a learnable balanced mixing mechanism to enhance robustness against scenario imbalance and distribution shift. Extensive experiments conducted on two public datasets demonstrate that the proposed method significantly outperforms baseline models in terms of overall performance and stability.

Keywords

Recommendation Systems, Multi-Scenario Recommendation, Multi-Task Prediction, E-Commerce, Attention Mechanism

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着电商平台业务形态的持续演进，现代推荐[1]与检索系统正由面向单场景、单目标的局部优化，逐步转向同时覆盖多场景与多任务的统一建模范式[2]。例如，在电商与内容推荐场景中，用户可能在不同入口(如首页、垂直页与搜索页)下产生点击、转化、停留等多源异构的行为信号。如何在同一模型中兼顾跨场景的语义差异与跨任务的目标关联，实现有效的表示学习与协同优化，已成为学术界与工业界共同关注的关键问题[3]。

多场景推荐[4]与多任务学习通常被视为缓解数据稀疏、提升泛化能力的重要途径。既有研究表明，不同场景之间往往共享用户与物品的潜在兴趣结构，而不同任务则从互补视角刻画用户偏好；因此，适度的信息共享能够促进知识迁移并改善整体表现[5]。然而，在真实电商系统中，不当的共享机制可能导致负迁移：部分任务或场景性能提升的同时，另一些任务或场景出现退化。该现象在多任务学习中常被称为跷跷板效应，在多场景学习中亦存在相近的表现形式。

为缓解上述冲突并实现更精细的共享控制，MoE 及其变体成为广泛采用的建模框架。典型方法如MMoE [6]通过共享专家网络并为各任务配置独立门控，实现任务级的信息选择与动态组合；Star [7]采用星型拓扑结构，显式区分场景共享表征与场景特有表征；PLE [8]进一步引入分层共享与任务特化专家，通过逐层抽取与解耦，降低单层共享带来的负迁移风险。总体而言，这类方法在多任务与电商多场景设

置下均展现出良好的有效性与可扩展性。

在更为复杂的电商多场景与多任务联合建模设定下, 相关研究进一步聚焦于层次化专家结构与自动化专家选择机制[9]。代表性工作 AESM² [10]通过对多场景层与多任务层进行层级堆叠, 并引入基于 KL 散度的专家选择策略, 在统一框架内同时刻画场景间与任务间关系, 且已在大规模工业系统中实现落地部署。近期的 M3oE [11]从多场景多任务视角出发, 提出多视角专家与自适应融合机制, 系统揭示了多场景多任务环境中跷跷板效应的普遍性。与此同时, PEPNet [12]等方法尝试借助个性化门控与参数调制, 在嵌入层与塔网络层注入用户与场景先验, 从而缓解多场景多任务学习中的偏置与分布差异问题。

尽管上述方法在结构设计与信息共享机制上持续演进, 现有研究仍存在两点关键局限。其一, 任务与电商场景的共享粒度缺乏语义层级上的自适应能力。多数方法仍主要依赖单层门控或静态专家分组来决定共享与特化关系, 即在同一语义层级上完成对所有任务与场景的知识分配[13]。然而, 在真实电商系统中, 不同任务与场景往往在低层表示上高度相关, 而在高层决策表示上差异显著。单层共享机制难以刻画这种随语义层级变化的相关结构, 从而限制了模型在复杂多场景多任务场景中的表示能力与迁移效果[14]。其二, 电商多场景多任务模型对特征交互的显式建模仍相对不足。尽管特征交互已被广泛证明是提升推荐性能的关键因素, 现有多场景多任务模型往往依赖 MLP 对拼接后的嵌入进行隐式拟合, 对字段-字段或维度-维度的高阶依赖刻画能力有限。近期的 MMFI [2]等工作开始关注场景-任务感知的特征交互重要性, 但其与 MoE 式共享结构的深度融合仍有待进一步探索与完善。

多场景多任务推荐同时面临三类耦合难点: 其一, 输入侧存在复杂的高阶特征交互, 仅依赖 MLP 往往难以充分刻画字段间的全局依赖; 其二, 不同场景之间存在显著的分布偏移, 需要在共享知识与场景特有偏置之间取得平衡; 其三, 不同任务在优化目标上可能存在不一致甚至冲突, 若共享机制控制不当, 容易产生负迁移与跷跷板效应。针对上述三类问题, 本文并非简单叠加已有模块, 而是构建了一条具有明确功能分工的递进式建模链路: 首先通过自注意力模块在输入侧显式增强高阶特征交互, 提升底层共享表示的判别性; 随后借助 Star 结构在统一参数底座上注入场景条件化偏置, 实现跨场景共性建模与场景差异适配; 最后采用 PLE 替代 MMoE 结构, 使共享信息与任务特有信息在不同语义层级中渐进式抽取与分离。由此, 形成了输入交互增强-场景对齐-任务解耦的协同框架, 更适合处理电商多场景多任务预测中的异质性与负迁移问题。

基于上述观察, 本文提出一种面向电商多场景多任务预测的分层专家模型(Self-Attentive Progressive Expert Network) (SAPENet)。该模型在学习多场景共享表示的基础上, 引入自注意力机制以显式刻画特征交互关系, 并采用 PLE 实现跨任务的层级化共享与特化, 从而在不同语义深度上自适应平衡共享知识与任务/场景特有信息。本文的主要贡献概括如下:

(1) 提出分层专家结构以缓解多任务与多场景学习中的负迁移问题。具体而言, 将共享-特化的归纳偏置由传统单层专家混合扩展为多层渐进抽取, 使跨任务知识迁移能够在不同语义层级自适应发生, 从结构上抑制负迁移并提升模型表示能力。

(2) 在嵌入层之后引入自注意力模块, 显式建模高阶特征依赖; 同时通过形状自适应的结构设计增强模型对不同特征组织方式的鲁棒性与可扩展性。

(3) 在两个公开数据集上开展了广泛实验, 结果表明所提出模型在整体性能与稳定性方面均取得显著改进, 从而验证了其有效性。

2. 电商多场景多任务推荐问题描述与定义

在电商多场景多任务推荐预测中, 设 U 和 I 表示用户集和项目集。 T 预测推荐任务旨在寻找一种同时优化 S 场景上的多场景多任务推荐任务的解决方案。对于场景 $s \in \{0, \dots, S\}$, 其中对于两个不同的场景,

可能存在一些共同的用户或者项目(也称为重叠),这些重叠的用户允许场景间共享知识,并达到比单独处理每个场景更好的推荐准确性。我们定义了一个全面的特征输入 x_s ,它由用户、项目和上下文信息组成。在我们的设置中,我们考虑所有预测任务(例如,点击率)的每个目标标签是 $\{0, 1\}$ 中的二进制信号,其定义用户是否提供正反馈(例如,点击)。即若标签为 1,表示用户与项目发生过交互(如点击、购买等行为),若标签为 0 则表示未观测到交互。然后,我们的研究目标是为每个任务 $t \in \{0, \dots, T\}$ 和场景 $s \in \{0, \dots, S\}$ 学习一个预测函数 $\hat{y}_{s,t} = f^{s,t}(x_s)$ 。

3. 面向电商的 SAPENet 模型算法实现

我们提出一种面向电商多场景多任务推荐的统一框架:在 Star 场景融合底座上,引入自注意力机制增强特征交互,并以 PLE 结构替代传统 MMoE,实现分层渐进的共享-任务特化抽取,从而显著缓解任务冲突带来的负迁移;同时保留场景/任务专家以提升在场景不均衡与分布漂移下的鲁棒性。总体框架结构如图 1 所示。

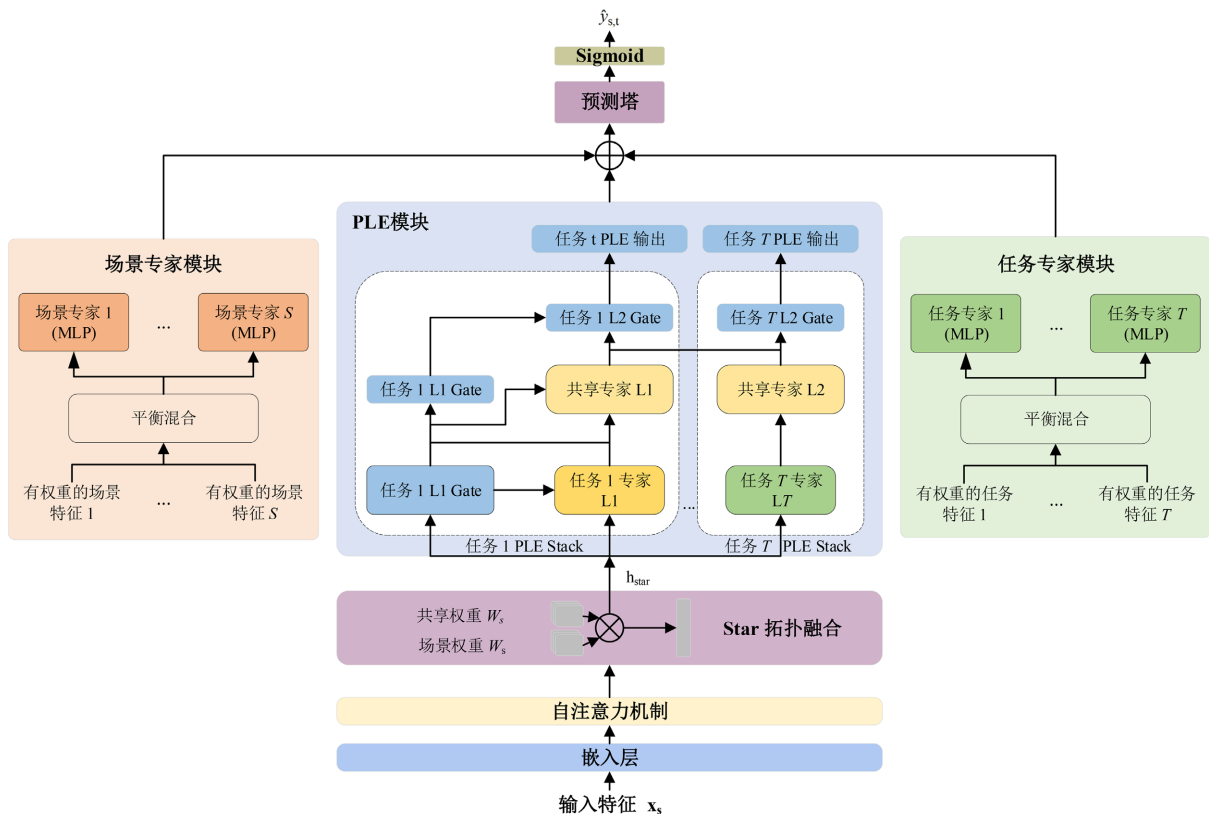


Figure 1. Framework diagram of the SAPENet model

图 1. SAPENet 模型框架图

3.1. 嵌入与自注意力增强表示

在定义了一个全面的特征输入 x_s 后,其中它包括用户、项目和上下文信息。首先将离散特征嵌入并拼接得到

$$e = \text{Embed}(x_s) \in \mathbb{R}^m \quad (1)$$

其中 m 为所有特征嵌入维度之和。通过引入自注意力模块以显式建模特征交互,得到注意力增强表示:

$$\tilde{e} = \text{Attn}(e) \quad (2)$$

3.2. Star 拓扑融合

Star 模块的目标是：在共享参数的基础上，为不同场景注入轻量场景偏置，并保留一条跳连以降低过拟合与梯度消失风险。

3.2.1. 场景条件化线性融合

对场景 s ，所实现的核心是共享权重 W_s 与场景权重 W_s 的逐元素相乘，再与输入表示做线性变换

$$z_s = \tilde{e}(W_s \odot W_s) + b_s + b_s \quad (3)$$

其中， $W_s \in \mathbb{R}^{m \times h}$ ， $W_s \in \mathbb{R}^{m \times h}$ ，输出维度为 h 。

3.2.2. 非线性投影与跳连

模型采用两层结构：一条跳连分支 $s' = \text{MLP}_{\text{skip}}(\tilde{e})$ ，以及主干 MLP:

$$h_* = \text{MLP}_*(z_s) + s' \quad (4)$$

其中 $h_* \in \mathbb{R}^p$ 作为后续多任务学习的共享底座表示， p 表示 Star 模块输出向量的维度。

3.3. PLE 结构

M3oE 模型采用 MMoE 结构，即对每个场景 - 任务对 (s, t) 使用一个门控单元，在共享专家输出上做加权组合。该结构能够同时建模多场景与多任务关系，但其共享信息提取主要依赖单层 MoE 式路由，难以显式表达底层更适合共享、高层更需要分化的层次性。同时，输入侧特征交互仍主要由 MLP 隐式拟合，对复杂高阶依赖的刻画有限。基于此，SAPENet 在保留 M3oE 场景专家、任务专家、平衡混合机制以及场景—任务塔网络的前提下，进行了两项针对性改进：一是在嵌入层后加入自注意力模块，以增强输入表示的高阶交互建模能力；二是以 PLE 替代原有单层共享专家路由，使共享与特化关系能够在多层语义空间中渐进式形成。因而，SAPENet 可以视为对 M3oE 共享表征学习机制的增强，而非对其整体框架的完全重构。而我们的模型的核心变化是用 PLE 取代 MMoE，使共享/任务特化在多个层级渐进抽取，从而更强地抑制任务间负迁移。

3.3.1. 第 l 层的专家集合

设共有 L 个 PLE 层，每层专家数为 M 。在第 l 层：

- 1) 共享专家： $\{E_{s,1}^{(l)}, \dots, E_{s,M}^{(l)}\}$
- 2) 任务 t 的特化专家： $\{E_{t,1}^{(l)}, \dots, E_{t,M}^{(l)}\}$

每个专家都是一个 MLP，输入维度随层推进更新。

3.3.2. 门控单元与层内混合

对任务 t 在层 l ，门控单元输出对 $2M$ 个专家(共享 M +任务特化 M)的权重：

$$\pi_t^{(l)} = \text{softmax}\left(W_{g,t}^{(l)} h_t^{(l-1)} + b_{g,t}^{(l)}\right) \in \mathbb{R}^{2M} \quad (5)$$

记第 k 个专家输出为 $u_{t,k}^{(l)} \in \mathbb{R}^{p^l}$ ，则该层输出为加权和：

$$h_t^{(l)} = \sum_{k=1}^{2M} \pi_{t,k}^{(l)} u_{t,k}^{(l)}, h_t^{(0)} = h_* \quad (6)$$

最终得到每个任务的 PLE 表示：

$$\mathbf{h}_t^{\text{PLE}} = \mathbf{h}_t^{(L)} \quad (7)$$

对比原 MMoE, 即 M3oE 的门控单元是针对每个 (s, t) 对共享专家集合加权(场景数目*任务数目个门控单元), 输出再与场景/任务专家项相加。而我们的模型是先得到仅按任务的 $\mathbf{h}_t^{\text{PLE}}$, 再在下一步与场景/任务专家增强项融合。

3.4. 电商场景/任务专家的平衡混合

为进一步缓解场景偏移与任务异质性, 我们保留了 M3oE 模型的场景专家与任务专家模块, 并用可学习标量进行平衡混合。

3.4.1. 场景专家与任务专家输出

$$\mathbf{g}_s = \mathbf{G}_s(\mathbf{h}_*), s = 0, \dots, S \quad (8)$$

$$\mathbf{q}_t = \mathbf{Q}_t(\mathbf{h}_*), t = 0, \dots, T \quad (9)$$

其中 $\mathbf{G}_s, \mathbf{Q}_t$ 为 MLP。

3.4.2. 平衡混合

对于每个场景 s , 用标量 $\lambda_s \in (0, 1)$ 控制本场景和其他场景均值的权重:

$$\bar{\mathbf{g}}_s = \lambda_s \mathbf{g}_s + \frac{1 - \lambda_s}{S - 1} \sum_{j \neq s} \mathbf{g}_j \quad (10)$$

同理对于任务 t :

$$\bar{\mathbf{q}}_t = \lambda_t \mathbf{q}_t + \frac{1 - \lambda_t}{T - 1} \sum_{k \neq t} \mathbf{q}_k \quad (11)$$

这样设计的作用是当某场景/任务数据稀疏或噪声大时, 模型可自适应引入其他场景/任务的稳定先验。

3.5. 融合表示与场景 - 任务塔预测

3.5.1. 融合表示

对每个场景 - 任务对 (s, t) , 最终构造融合特征为:

$$\mathbf{f}_{s,t} = \mathbf{h}_t^{\text{PLE}} + \alpha_s \bar{\mathbf{g}}_s + \alpha_t \bar{\mathbf{q}}_t \quad (12)$$

其中, $\alpha_s = \omega_{\text{exp}_s}$, $\alpha_t = \omega_{\text{exp}_t}$ 为可学习标量。

3.5.2. 场景 - 任务塔预测

每个 (s, t) 配置一个塔网络 $\text{Tower}_{s,t}(\cdot)$, 输出可表示为:

$$\hat{y}_{s,t} = \sigma(\text{Tower}_{s,t}(\mathbf{f}_{s,t})), \sigma(z) = \frac{1}{1 + e^{-z}} \quad (13)$$

推理时按样本所属场景 s 选择对应输出, 得到最终的任务预测 $\hat{y}_{s,t}$ 。从信息流角度看, SAPENet 的三个核心组件对应于不同层次的信息处理阶段。自注意力模块首先在嵌入空间中对字段级表示进行重加权, 使模型在进入场景建模之前即可显式捕获用户、物品与上下文之间的高阶依赖, 从而提升基础表示质量。Star 模块随后在共享参数底座上引入场景条件化调制, 使不同场景样本先在统一语义空间内完成对齐, 再将场景差异以轻量偏置的方式保留下来。PLE 则进一步对经过场景融合后的表示执行逐层共享与逐层特化: 底层优先保留可迁移的通用模式, 高层逐步聚焦任务相关的判别信息。因而, 整个模型的信息流

并非一次性共享或一次性分裂，而是沿着全局交互 - 场景适配 - 任务分化的路径递进展开。

3.6. 训练目标与损失函数

我们选择二元交叉熵作为损失函数，

$$L = \sum_s^{|\mathcal{S}|} \sum_t^{|\mathcal{T}|} \text{BCE}(\hat{y}_{s,t}, y_{s,t}) \quad (14)$$

4. 实验与分析

为了充分评估所提出的 SAPENet 模型性能，我们在电商平台公开的两个真实数据集 AliCCP 和 KuaiSAR [15] 上进行了对比实验分析。实验环境基于 Windows 10 64 位操作系统，CPU 是 Intel(R) Core(TM) i7-9700F，GPU 为 NVIDIA GeForce GTX 1660，显卡拥有 9.9 GB 的显存。

4.1. 数据集

本节用了两个数据集进行了实验：一个公共数据集 AliCCP 数据集(<https://tianchi.aliyun.com/dataset/408>) 和一个来自于快手 APP 上用户真实消费情况的真实工业 KuaiSAR 数据集(<https://ethan00si.github.io/KuaiSAR/>)，其中划分为搜索和推荐场景。AliCCP 数据集有 3 个场景，任务划分成点击和购买，按照 8:1:1 的比例划分为训练集、验证集和测试集。KuaiSAR 数据集划分为搜索和推荐 2 个场景，其中任务划分为点击、喜欢、分享，数据集的统计结果如表 1 所示。

Table 1. Statistical results of the dataset

表 1. 数据集的统计结果

数据集	用户数量	场景数量	任务数量	样本数量
AliCCP	158,565	3	2	85,316,519
KuaiSAR	39,469	2	3	14,524,508

4.2. 评价指标

我们选用 AUC 与 LogLoss 作为主要评估指标。所有实验结果均报告为在 5 个不同随机种子下重复运行所得的平均值，以降低随机性带来的影响并提高结论的可靠性。

4.3. 实验设置

为保证比较的公平性，我们在验证集上进行超参数搜索以确定最优配置，并在测试集上报告各模型的最终性能。嵌入维度设置为 16。对于 AliCCP 数据集，学习率设为 $1e-2$ ，批处理大小设为 32768；对于 KuaiSAR 数据集，学习率设为 $1e-4$ ，批处理大小设为 512。在模型结构方面，我们统一采用两层 PLE 模块。

4.4. 结果分析

为全面评估模型性能，我们选取了具有代表性的先进方法作为对比基线。为确保比较的系统性，我们从四种不同设置出发进行对比：单场景单任务、单场景多任务、多场景单任务以及多场景多任务推荐模型。实验结果中，最优结果以加粗标示，次优结果以下划线标示。结果如表 2，表 3 所示。

下表 2 给出了 AliCCP 数据集上的模型对比结果。可以观察到，在 d1,t0 列中，所提出方法相较于原始 M3oE 在 AUC 上提升 3.85%；在 d0,t1 与 d1,t1 两列中，分别提升 1.5% 与 1.12%。总体而言，该结果表明：将 M3oE 中原有的 MMoE 结构替换为 PLE 结构能够更有效地缓解异构场景与异构任务之间的负迁

移问题。进一步地，引入自注意力机制后，专家表征的判别性得到增强，尤其在稀疏或噪声较高的目标上表现更为显著。与此同时，我们的方法在多个设置下取得了持续更低的 LogLoss，说明该模型不仅提升了排序识别能力，还能够输出更具校准性的概率估计。

Table 2. The experimental results of the AliCCP dataset. d0 and t0 represent the AUC results of scenario 0 and task 0 respectively. The best results are marked in bold, and the subbest results are marked with underlines. *Indicates a significance level p value < 0.05 relative to the optimal baseline. All results are the averages of five repeated runs using different random seeds
表 2. AliCCP 数据集实验结果。d0,t0 表示场景 0 和任务 0 的 AUC 结果，最优的结果用加粗标出，次优的结果用下划线标出。*表示相对于最佳基线的显著性水平 p 值<0.05。所有结果都是使用不同随机种子的 5 次重复运行的平均值

	d0,t0	d0,t1	d1,t0	d1,t1	d2,t0	d2,t1	All AUC	Logloss
Single-task	0.5503	0.5141	0.567	0.5549	0.553	0.5043	<u>0.5316</u>	0.0023
MMoE	0.5451	0.5049	0.5723	0.5528	0.5399	0.5021	0.5244	0.0026
PLE	0.5479	0.5088	<u>0.5729</u>	0.5591	0.5477	0.504	0.5216	0.0024
Star	<u>0.5536</u>	<u>0.5201</u>	0.5526	0.5519	0.5462	0.5033	0.5219	0.0019
SarNet	0.5488	0.5091	0.5712	0.5571	0.5511	0.5013	0.5269	0.002
M2M	0.5501	0.5122	0.5614	0.5588	0.546	<u>0.508</u>	0.5296	0.0016
M3oE	0.5512	0.5125	0.5528	<u>0.5594</u>	0.5475	0.5093	0.5344	0.0014
SAPENet	0.5551*	0.5275*	0.5913*	0.5706*	<u>0.5521</u>	0.505	0.5266	0.0013

此外，从整体对比可以发现，传统的单场景模型在 Aliccp 数据集上的表现明显弱于多任务模型与多场景模型，这表明简单的单场景建模难以有效适配真实系统中并存的多个场景。进一步地，多场景模型相较于仅进行多任务学习的模型通常取得更优结果，说明仅依赖任务共享机制不足以刻画复杂的跨场景差异与共性结构，从而凸显了专门的多场景建模策略的必要性。最后，在多场景背景下联合进行多任务优化往往优于仅进行单任务预测，这主要得益于模型能够在场景 - 任务二维空间内更细粒度地建模关联关系，从而提升预测精度与整体泛化能力。

表 3 给出了模型在 KuaiSAR 数据集上的对比结果。可以观察到，改进后的模型相较于原始模型在各项指标上均取得了一致性提升。这一性能增益主要归因于两个方面。

Table 3. The experimental results of the KuaiSAR dataset
表 3. KuaiSAR 数据集实验结果

	d0,t0	d0,t1	d0,t2	d1,t0	d1,t1	d1,t2	All AUC	Logloss
Single-task	0.9801	0.9001	0.8479	<u>0.9419</u>	0.8724	0.8203	0.9121	0.0934
MMoE	0.9836	0.901	0.8533	0.9377	0.8744	0.8199	0.9157	0.0939
PLE	0.9839	0.9021	0.8512	0.931	0.8759	0.8192	<u>0.9206</u>	0.0936
Star	<u>0.9856</u>	0.9013	<u>0.8594</u>	0.9355	0.8706	0.8223	0.9101	0.0935
SarNet	0.9816	<u>0.904</u>	0.851	0.9306	0.8711	0.8269	0.9112	0.0933
M2M	0.9813	0.9016	0.8569	0.9411	0.8771	0.8216	0.9202	0.0925
M3oE	0.9843	0.9002	0.8488	0.9392	<u>0.8778</u>	<u>0.828</u>	0.9113	0.0921
SAPENet	0.9872*	0.9079*	0.8669*	0.9603*	0.8840*	0.83*	0.9209	0.0897*

首先, PLE 将表示学习过程显式划分为全局共享信息、场景相关信息与任务特有信息三个层级, 从结构上对信息流进行约束。

这种分层解耦机制有效减少了无关任务梯度对目标任务预测分布的干扰, 缓解了多任务优化中的梯度冲突问题。其次, 自注意力机制对专家输出进行动态加权, 使模型能够在样本层面自适应地放大高置信专家、抑制无关或噪声专家。本质上, 这种机制降低了预测分布的方差, 使输出概率在不同样本之间更加平滑与稳定, 从而提升整体泛化能力。在 AliCCP 的不同子场景以及 KuaiSAR 的搜索/推荐双场景中, 用户意图分布与行为转化路径存在显著差异。在此背景下, PLE 通过层级结构控制信息流动顺序: 首先在底层提取可共享但弱绑定的通用表示, 随后逐层引入场景特定与任务特定专家进行细化建模。这种渐进式解耦策略避免了低相关任务在早期阶段的过度共享, 使跨任务交互更加可控。

总体而言, 在复杂且分布差异显著的 KuaiSAR 数据集上所获得的一致性性能提升, 验证了该模型在真实多场景多任务推荐系统中的鲁棒性与可扩展性。PLE 与自注意力机制的结合不仅减少了异构任务之间的梯度干扰, 还能够动态抑制噪声专家的影响, 从而输出更加稳定且校准良好的概率估计, 这一点亦从改进模型在 LogLoss 上的持续优势中得到体现。

4.5. 消融实验

为了进一步研究改进模型的有效性, 我们通过单独移除关键部件进行了消融研究。具体地说, 分别移除自注意力机制(w/o 自注意力机制)和 PLE 结构(w/o PLE 结构), 设计这两个变体模型和我们所提出的模型进行比较, 结果如表 4, 表 5 所示。

Table 4. Results of ablation experiments on the AliCCP dataset. The result is the average of five single tests

表 4. AliCCP 数据集消融实验结果。结果是 5 次单次测试的平均值

	d0,t0	d0,t1	d1,t0	d1,t1	d2,t0	d2,t1	All AUC	Logloss
w/o 自注意力机制	<u>0.5477</u>	0.5243	<u>0.5889</u>	0.5448	<u>0.5434</u>	0.5282	<u>0.5279</u>	0.0088
w/o PLE 结构	0.526	0.558	0.5684	0.5825	0.5222	0.5031	0.5296	0.0016
SAPENet	0.5551	<u>0.5275</u>	0.5913	<u>0.5706</u>	0.5521	<u>0.505</u>	0.5266	0.0013

Table 5. Results of ablation experiments on the KuaiSAR dataset

表 5. KuaiSAR 数据集消融实验结果

	d0,t0	d0,t1	d0,t2	d1,t0	d1,t1	d1,t2	All AUC	Logloss
w/o 自注意力机制	0.9844	0.8963	0.8424	0.9484	0.8716	<u>0.7979</u>	0.908	0.0929
w/o PLE 结构	<u>0.9866</u>	<u>0.9033</u>	<u>0.8571</u>	<u>0.9573</u>	<u>0.8807</u>	0.7968	<u>0.9158</u>	0.0908
SAPENet	0.9872	0.9079	0.8669	0.9603	0.8840	0.83	0.9209	0.0897

在两个数据集上, 当移除自注意力机制时, LogLoss 均出现明显上升, 表明该模块对提升概率预测的校准性具有重要作用。从排序性能来看, 去除自注意力后, 各场景 - 任务组合下的 AUC 值普遍低于完整模型, 进一步验证了输入级自注意力模块的有效性。具体而言, 该模块在进入多场景融合结构之前显式建模特征之间的全局依赖关系, 使模型能够学习到更加鲁棒且对场景变化不敏感的表达, 从而为后续的场景融合结构(如 Star)与多任务优化提供更高质量的共享特征基础。对于 PLE 结构的消融, 虽然在个别任务上观察到局部提升(例如 AliCCP 数据集中的 d1,t1 列), 但同时会导致其他任务出现显著退化, 呈现出典型的以一损多现象。相比之下, 完整模型在所有场景 - 任务组合上表现更加均衡, 没有出现明显的

性能偏置。这一现象说明,传统共享专家机制在面对任务差异较大的情形时,容易产生表示冲突与梯度干扰。PLE 通过逐层共享、逐层解耦的结构设计,使底层网络充分学习通用表示,而高层逐步聚焦于场景或任务特定模式,从而在结构层面缓解负迁移问题。

综上,消融实验充分验证了本文两个核心设计的有效性:输入级自注意力模块增强了跨场景表征能力与预测稳定性,而 PLE 结构通过分层共享与任务解耦有效抑制了多任务负迁移。二者的协同作用使模型在不同数据集、不同场景与不同任务设置下均表现出稳定且一致的性能提升。

5. 结束语

本研究提出一种融合注意力增强表示学习与 PLE 分层专家机制的电商多场景多任务统一推荐框架 SAPENet 模型,旨在有效缓解跨任务与跨场景建模中的负迁移问题。具体而言,我们在多场景多任务推荐体系中引入输入级自注意力模块,以显式刻画特征之间的全局依赖关系,从而学习更具跨场景鲁棒性的表示。针对传统 MMoE 在任务差异较大时易产生表示冲突与负迁移的不足,我们采用分层 PLE 结构替代共享专家门控机制,通过共享专家与任务特定专家的渐进式解耦实现更稳定的跨任务知识迁移。大量实验结果表明,所提出模型在多个真实世界数据集上显著优于现有代表性方法,验证了其有效性与鲁棒性。

基金项目

国家重点研发计划项目(2023YFB3308802),国家自然科学基金项目(52275480),贵州省自然科学基金(QKH MS [2025] 601, QKH QN [2024] 160)。

参考文献

- [1] 吕学强,王夏雨,马登豪. 面向推荐系统的用户兴趣建模综述[J]. 计算机工程与应用, 2025, 61(21): 15-29.
- [2] Song, D., Yang, E., Guo, G., Shen, L., Jiang, L. and Wang, X. (2024) Multi-Scenario and Multi-Task Aware Feature Interaction for Recommendation System. *ACM Transactions on Knowledge Discovery from Data*, **18**, 1-20. <https://doi.org/10.1145/3651312>
- [3] 任衍栋,张东,李冠宇. 融合注意力与结构降噪的对比学习知识感知推荐[J]. 计算机工程与应用, 2025, 61(17): 232-240.
- [4] Ning, W., Yan, X., Liu, W., Cheng, R., Zhang, R. and Tang, B. (2023) Multi-Domain Recommendation with Embedding Disentangling and Domain Alignment. <http://arxiv.org/abs/2308.05508>
- [5] Zhu, J., Wang, Y., Lin, J., Qin, J., Tang, R., Zhang, W., et al. (2024) M-Scan: A Multi-Scenario Causal-Driven Adaptive Network for Recommendation. *Proceedings of the ACM Web Conference 2024*, Singapore, 13-17 May 2024, 3844-3853. <https://doi.org/10.1145/3589334.3645635>
- [6] Ma, J., Zhao, Z., Yi, X., Chen, J., Hong, L. and Chi, E.H. (2018) Modeling Task Relationships in Multi-Task Learning with Multi-Gate Mixture-of-Experts. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, 19-23 August 2018, 1930-1939. <https://doi.org/10.1145/3219819.3220007>
- [7] Sheng, X.R., Zhao, L., Zhou, G., et al. (2021) One Model to Serve All: Star Topology Adaptive Recommender for Multi-Domain CTR Prediction. <http://arxiv.org/abs/2101.11427>
- [8] Tang, H., Liu, J., Zhao, M. and Gong, X. (2020) Progressive Layered Extraction (PLE): A Novel Multi-Task Learning (MTL) Model for Personalized Recommendations. *Fourteenth ACM Conference on Recommender Systems*, Brazil, 22-26 September 2020, 269-278. <https://doi.org/10.1145/3383313.3412236>
- [9] Ouyang, W., Zhang, X., Zhao, L., et al. (2020) MiNet: Mixed Interest Network for Cross-Domain Click-Through Rate Prediction. <http://arxiv.org/abs/2008.02974>
- [10] Zou, X., Hu, Z., Zhao, Y., et al. (2022) Automatic Expert Selection for Multi-Scenario and Multi-Task Search. <http://arxiv.org/abs/2205.14321>
- [11] Zhang, Z., Liu, S., Yu, J., Cai, Q., Zhao, X., Zhang, C., et al. (2024) M3oE: Multi-Domain Multi-Task Mixture-of-Experts Recommendation Framework. *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Washington DC, 14-18 July 2024, 893-902.

- <https://doi.org/10.1145/3626772.3657686>
- [12] Chang, J., Zhang, C., Hui, Y., Leng, D., Niu, Y., Song, Y., *et al.* (2023) PEPNet: Parameter and Embedding Personalized Network for Infusing with Personalized Prior Information. *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Long Beach, 6-10 August 2023, 3795-3804. <https://doi.org/10.1145/3580305.3599884>
- [13] Shazeer, N., Mirhoseini, A., Maziarz, K., *et al.* (2017) Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer. <http://arxiv.org/abs/1701.06538>
- [14] Schoenauer-Sebag, A., Heinrich, L., Schoenauer, M., *et al.* (2019) Multi-Domain Adversarial Learning. <http://arxiv.org/abs/1903.09239>
- [15] Sun, Z., Si, Z., Zang, X., Leng, D., Niu, Y., Song, Y., *et al.* (2023) KuaiSAR: A Unified Search and Recommendation Dataset. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, Birmingham, 21-25 October 2023, 5407-5411. <https://doi.org/10.1145/3583780.3615123>